

## **Remarks**

In the present response, no claims are amended; claim 5 is canceled; and four claims (15-18) are newly presented. Claims 1-4 and 6-18 are presented for examination. Applicant believes that no new matter is entered.

### **I. Claim Rejections: 35 USC § 102**

Claims 1, 5-6, and 10 are rejected under 35 USC § 102 as being anticipated by an article entitled “Mercator: A scalable, extensive Web Crawler” by Heydon et al. (hereinafter Heydon). This rejection is traversed.

A proper rejection of a claim under 35 U.S.C. §102 requires that a single prior art reference disclose each element of the claim. See MPEP § 2131, also, *W.L. Gore & Assoc., Inc. v. Garlock, Inc.*, 721 F.2d 1540, 220 U.S.P.Q. 303, 313 (Fed. Cir. 1983). Since Heydon neither teaches nor suggests each element in claims 1, 6, and 10, these claims are allowable over Heydon. The rejection is moot regarding claim 5 since this claim is canceled.

#### **Claim 1**

Independent claim 1 recites numerous limitations that are not taught or suggested in Heydon. For example, claim 1 recites “a plurality of web crawlers.” By contrast, Heydon does not teach or suggest a plurality of web crawlers. Heydon teaches a single web crawler (see Abstract: “This paper describes Mercator, a scalable, extensible web crawler ....”). Applicant admits that Section 3.1 paragraph 1 states: “Crawling is performed by multiple worker threads.” Multiple worker threads, though, are not a plurality of web crawlers. In fact, Fig. 1 of Heydon teaches a single web crawler.

As another example, claim 1 recites “assigning a web crawler identifier to each one of the plurality of web crawlers.” Heydon does not teach or suggest this limitation. The Office Action cites Section 3.2, third paragraph of Heydon for teaching this limitation. This section of Heydon teaches that Mercators’ URL frontier includes distinct FIFO subqueues; one FIFO subqueue per worker thread. This section further states: “Second, when a new URL is added, the FIFO subqueue in which it is placed is

determined by the URL's canonical host name." No where does this section teach or suggest assigning an identifier to each one of a plurality of web crawlers.

As another example, claim 1 recites:

determining a web crawler identifier to which the representation corresponds; and

when the determined web crawler identifier is not assigned to the respective web crawler, sending the identified address to the web crawler to which the determined web crawler identifier is assigned.

Heydon does not teach or suggest these limitations. The Office Action repeatedly cites Section 3.2. This section of Heydon teaches a data structure (URL frontier) that contains all the URLs that remain to be downloaded within a single web crawler. The claimed limitations in claim 1, though, are not shown or suggested.

#### **Claim 6**

Independent claim 6 recites numerous limitations that are not taught or suggested in Heydon. For example, claim 6 recites "a plurality of web crawlers." By contrast, Heydon does not teach or suggest a plurality of web crawlers. Heydon teaches a single web crawler (see Abstract: "This paper describes Mercator, a scalable, extensible web crawler ..."). Applicant admits that Section 3.1 paragraph 1 states: "Crawling is performed by multiple worker threads." Multiple worker threads, though, are not a plurality of web crawlers. In fact, Fig. 1 of Heydon teaches a single web crawler.

As another example, claim 6 recites "wherein each web crawler has been assigned a web crawler identifier." Heydon does not teach or suggest this limitation. The Office Action cites Section 3.2, third paragraph of Heydon for teaching this limitation. This section of Heydon teaches that Mercators' URL frontier includes distinct FIFO subqueues; one FIFO subqueue per worker thread. This section further states: "Second, when a new URL is added, the FIFO subqueue in which it is placed is determined by the URL's canonical host name." No where does this section teach or suggest a plurality of web crawlers with each web crawler assigned a web crawler identifier.

As another example, claim 6 recites:

for each respective web crawler:  
a main web crawler module ...  
determining a web crawler identifier to which the  
representation corresponds; and  
when the determined web crawler identifier is not  
assigned to the respective web crawler, sending the  
identified address to a destination web crawler comprising  
the web crawler to which the determined web crawler  
identifier is assigned.

Heydon does not teach or suggest these limitations. The Office Action repeatedly cites Section 3.2. This section of Heydon teaches a data structure (URL frontier) that contains all the URLs that remain to be downloaded within a single web crawler. The claimed limitations in claim 6, though, are not shown or suggested.

#### **Claim 10**

Independent claim 10 recites numerous limitations that are not taught or suggested in Heydon. For example, claim 10 recites:

determining a web crawler identifier to which the  
representation corresponds; and  
when the determined web crawler identifier is not  
assigned to the respective web crawler, sending the  
identified address to a destination web crawler comprising  
the web crawler to which the determined web crawler  
identifier is assigned.

Heydon does not teach or suggest these limitations. The Office Action repeatedly cites Section 3.2. This section of Heydon teaches a data structure (URL frontier) that contains all the URLs that remain to be downloaded within a single web crawler. The claimed limitations in claim 10, though, are not shown or suggested.

## II. Claim Rejections: 35 USC § 102(e)

Claims 1-14 are rejected under 35 USC § 102(e) as being anticipated by Najork et al. (USPN 6,377,984, hereinafter Najork). This rejection is traversed.

A proper rejection of a claim under 35 U.S.C. §102(e) requires that a single prior art reference disclose each element of the claim. See MPEP § 2131, also, *W.L. Gore & Assoc., Inc. v. Garlock, Inc.*, 721 F.2d 1540, 220 U.S.P.Q. 303, 313 (Fed. Cir. 1983). Since Najork neither teaches nor suggests each element in claims 1-4 and 6-14, these claims are allowable over Najork. The rejection is moot regarding claim 5 since this claim is canceled.

### Claim 1

Independent claim 1 recites numerous limitations that are not taught or suggested in Najork. For example, claim 1 recites “a plurality of web crawlers.” By contrast, Najork does not teach or suggest a plurality of web crawlers. Najork teaches a single web crawler:

FIG. 1 shows an exemplary embodiment of a distributed computer system 100. The distributed computer system 100 includes **a web crawler 102** connected to a network 103 through a network interconnection 110. (Col. 3, lines 60-63: emphasis added).

Applicant admits that Fig. 1 of Najork shows a web crawler 102 with memory 118 that includes threads 130. Multiple threads, though, are not a plurality of web crawlers. In fact, Fig. 1 of Najork teaches a single web crawler.

As another example, claim 1 recites “assigning a web crawler identifier to each one of the plurality of web crawlers.” Najork does not teach or suggest this limitation. The Office Action cites identifier “r” (Figs. 2-4) to teach one of a plurality of web crawlers (each thread being a crawler, see also Fig. 3B). These figures and accompanying description do not teach or suggest assigning an identifier to each one of a plurality of web crawlers. By contrast, the figures are generally directed to FIFO queues for a single web crawler.

As another example, claim 1 recites:

determining a web crawler identifier to which the representation corresponds; and

when the determined web crawler identifier is not assigned to the respective web crawler, sending the identified address to the web crawler to which the determined web crawler identifier is assigned.

Najork does not teach or suggest these limitations. The Office Action cites Steps 302-304, 508 and 306, 510, 554. These steps in Najork are generally directed to FIFO queues for URLs downloaded within a single web crawler. The claimed limitations in claim 1, though, are not shown or suggested.

Dependent claims 2-4 depend from claim 1 and thus inherit all the limitations of base claim 1. As such, claims 2-4 are also allowable over Najork. Further, these dependent claims contain numerous limitations not taught or suggested in Najork.

#### **Claim 6**

Independent claim 6 recites numerous limitations that are not taught or suggested in Najork. For example, claim 6 recites “a plurality of web crawlers.” By contrast, Najork does not teach or suggest a plurality of web crawlers. Najork teaches a single web crawler:

FIG. 1 shows an exemplary embodiment of a distributed computer system 100. The distributed computer system 100 includes **a web crawler 102** connected to a network 103 through a network interconnection 110. (Col. 3, lines 60-63: emphasis added).

Applicant admits that Fig. 1 of Najork shows a web crawler 102 with memory 118 that includes threads 130. Multiple threads, though, are not a plurality of web crawlers. In fact, Fig. 1 of Najork teaches a single web crawler.

As another example, claim 6 recites “wherein each web crawler has been assigned a web crawler identifier.” Najork does not teach or suggest this limitation. Najork does not teach or suggest this limitation. The Office Action cites identifier “r” (Figs. 2-4) to

each one of a plurality of web crawlers (each thread being a crawler, see also Fig. 3B). These figures and accompanying description do not teach or suggest assigning an identifier to teach one of a plurality of web crawlers. By contrast, the figures are generally directed to FIFO queues for a single web crawler.

As another example, claim 6 recites:

for each respective web crawler:  
a main web crawler module ...  
determining a web crawler identifier to which the representation corresponds; and  
when the determined web crawler identifier is not assigned to the respective web crawler, sending the identified address to a destination web crawler comprising the web crawler to which the determined web crawler identifier is assigned.

Najork does not teach or suggest these limitations. The Office Action cites Steps 302-304, 508 and 306, 510, 554. These steps in Najork are generally directed to FIFO queues for URLs downloaded within a single web crawler. The claimed limitations in claim 6, though, are not shown or suggested.

Dependent claims 7-9 depend from claim 6 and thus inherit all the limitations of base claim 6. As such, claims 7-9 are also allowable over Najork. Further, these dependent claims contain numerous limitations not taught or suggested in Najork.

### **Claim 10**

Independent claim 10 recites numerous limitations that are not taught or suggested in Najork. For example, claim 10 recites:

determining a web crawler identifier to which the representation corresponds; and  
when the determined web crawler identifier is not assigned to the respective web crawler, sending the identified address to a destination web crawler comprising the web crawler to which the determined web crawler identifier is assigned.

Najork does not teach or suggest these limitations. The Office Action cites Steps 302-304, 508 and 306, 510, 554. These steps in Najork are generally directed to FIFO queues for URLs downloaded within a single web crawler. The claimed limitations in claim 10 are not shown or suggested.

Dependent claims 11-14 depend from claim 10 and thus inherit all the limitations of base claim 10. As such, claims 11-14 are also allowable over Najork. Further, these dependent claims contain numerous limitations not taught or suggested in Najork.

### **III. Claim Rejections: 35 USC § 102(f)**

Claims 1-14 are rejected under 35 USC § 102(f) because applicant did not invent the claimed subject matter. This rejection is traversed.

In Section I of this response, Applicant has demonstrated that the claimed invention is patentable over Heydon. The rejection is moot regarding claim 5 since this claim is canceled.

### **IV. Claim Rejections: 35 USC § 102(e)**

Claims 1-14 are rejected under 35 USC § 102(e) as being anticipated by Eichstaedt et al. (USPN 6,182,085, hereinafter Eichstaedt). This rejection is traversed.

A proper rejection of a claim under 35 U.S.C. §102(e) requires that a single prior art reference disclose each element of the claim. See MPEP § 2131, also, *W.L. Gore & Assoc., Inc. v. Garlock, Inc.*, 721 F.2d 1540, 220 U.S.P.Q. 303, 313 (Fed. Cir. 1983). Since Eichstaedt neither teaches nor suggests each element in claims 1-4 and 6-14, these claims are allowable over Eichstaedt. The rejection is moot regarding claim 5 since this claim is canceled.

### **Claim 1**

Independent claim 1 recites numerous limitations that are not taught or suggested in Eichstaedt. For example, claim 1 recites “assigning a web crawler identifier to each one of the plurality of web crawlers.” Eichstaedt does not teach or suggest this limitation. The Office Action cites Col. 10 and gatherer processor id “i” as teaching this limitation.

This section of Eichstaedt teaches how to partition the web-graph among numerous gatherers or processors:

Assuming that one version of the present invention has k “gatherers” or processors, the web-graph is divided into k sub-graphs  $W_1, \dots, W_k$ . Each sub-graph is mapped to a processor (e.g.,  $W_i$  to processor i). (Col. 10, lines 18-21).

Thus, this section of Eichstaedt teaches how to divide the web-graph between processors. This section does not teach or suggest assigning a web crawler identifier to each processor or gatherer.

As another example, claim 1 recites downloading data sets and identifying addresses of one or more referred data sets. For each identified address, claim 1 specifically recites “generating a representation of the host computer identifier” and “determining a web crawler identifier to which the representation corresponds.” Eichstaedt does not teach this limitation. As noted, Eichstaedt does not assign web crawler identifiers to each gatherer or processor. As such, Eichstaedt does not generate a representation of a host computer identifier and then determine a web crawler identifier to which the representation corresponds.

The Office Action relies on Fig. 6 and Col. 6 (for example, lines 39-67 and 2-38). These sections in Eichstaedt teach dividing the web-space (the URL space) into sub-spaces and assigning sub-spaces to certain processors (Col. 6, lines 30-32). When new URLs are added to the web-space, the processor processes URLs belonging to its sub-space and routes other URLs (i.e., those not belonging to its sub-space) to the proper processor (Col. 6, lines 33-38). Notice though, that Eichstaedt does not generate a representation of a host computer identifier and then determine a web crawler identifier to which the representation corresponds.

Dependent claims 2-4 depend from claim 1 and thus inherit all the limitations of base claim 1. As such, claims 2-4 are also allowable over Eichstaedt. Further, these dependent claims contain numerous limitations not taught or suggested in Eichstaedt.

### **Claim 6**



Independent claim 6 recites numerous limitations that are not taught or suggested in Eichstaedt. For example, claim 6 recites “wherein each web crawler has been assigned a web crawler identifier.” For the reasons discussed above in connection with claim 1, Eichstaedt does not teach or suggest this limitation.

As another example, claim 6 recites a main web crawler module that identifies addresses of referred data sets, wherein each identified address includes a host computer identifier. An address distribution module processes the identified addresses and includes instructions for “generating a representation of the host computer identifier” and “determining a web crawler identifier to which the representation corresponds.” For the reasons discussed above in connection with claim 1, Eichstaedt does not teach or suggest this limitation.

Dependent claims 7-9 depend from claim 6 and thus inherit all the limitations of base claim 6. As such, claims 7-9 are also allowable over Eichstaedt. Further, these dependent claims contain numerous limitations not taught or suggested in Eichstaedt.

#### **Claim 10**

Independent claim 10 recites numerous limitations that are not taught or suggested in Eichstaedt. For example, claim 10 recites “wherein each web crawler has been assigned a web crawler identifier.” For the reasons discussed above in connection with claim 1, Eichstaedt does not teach or suggest this limitation.

As another example, claim 10 recites a main web crawler module that identifies addresses of referred data sets, wherein each identified address includes a host computer identifier. An address distribution module processes the identified addresses and includes instructions for “generating a representation of the host computer identifier” and “determining a web crawler identifier to which the representation corresponds.” For the reasons discussed above in connection with claim 1, Eichstaedt does not teach or suggest this limitation.

Dependent claims 11-14 depend from claim 10 and thus inherit all the limitations of base claim 10. As such, claims 11-14 are also allowable over Eichstaedt. Further, these dependent claims contain numerous limitations not taught or suggested in Eichstaedt.

## V. New Claims

Applicant submits new claims 15-18. These claims recite numerous limitations that are not taught or suggested, alone or in combination, by the art of record.

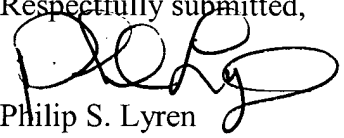
## CONCLUSION

In view of the above, Applicant believes claims 1-4 and 6-18 are in condition for allowance. Allowance of these claims is respectfully requested.

Any inquiry regarding this Amendment and Response should be directed to Philip S. Lyren at Telephone No. (281) 514-8236, Facsimile No. (281) 514-8332. In addition, all correspondence should continue to be directed to the following address:

**Hewlett-Packard Company**  
Intellectual Property Administration  
P.O. Box 272400  
Fort Collins, Colorado 80527-2400

Respectfully submitted,

  
Philip S. Lyren  
Reg. No. 40,709  
Ph: 281-514-8236

CERTIFICATE UNDER 37 C.F.R. 1.8: The undersigned hereby certifies that this paper or papers, as described herein, are being deposited in the United States Postal Service, as first class mail, in an envelope address to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on this 12th day of July, 2004.

By   
Name: Be Henry